NONPARAMETRIC ESTIMATION OF A HETEROGENEOUS DEMAND FUNCTION UNDER THE SLUTSKY RESTRICTION

by

Richard Blundell

Joel L. Horowitz

Matthias Parey

INTRODUCTION

- This talk is about nonparametric estimation of a demand function that is not additively separable.
- We illustrate the methods with an application to the demand for gasoline in the U.S.
- Economic theory does not provide a finite-dimensional parametric model of demand.
- Additive separability occurs only under restrictive assumptions about preferences.

MOTIVATION

- This motivates use of nonparametric methods and a non-separable specification to estimate dependence of demand on price and income.

- But a nonparametric estimate of the demand function is noisy due to random sampling errors.
    - The estimated function is wiggly and non-monotonic.
    - Some estimates of deadweight losses have incorrect signs and are, therefore, nonsensical.

POSSIBLE REMEDY

- Impose a parametric or semiparametric structure on the demand function.
- But there is no guarantee that such a structure is consistent with economic theory or otherwise correct or approximately correct.
- Demand estimation using a misspecified model can give seriously misleading results.

AN ALTERNATIVE APPROACH

- We impose structure by using a shape restriction from economic theory.
- Specifically, we impose the Slutsky restriction ofr consumer theory on an otherwise fully nonparametric estimate of the demand function.
- This yields will-behaved estimates of the demand function and deadweight losses.

ADVANTAGES OF THE APPROACH

- Maintains flexibility of nonparametric estimation.
- Is consistent with the theory of the consumer.
- Avoids using arbitrary and possibly incorrect parametric or semiparametric restrictions to stabilize estimates.
- Slutsky constrained nonparametric estimates reveal features of the demand function that are not present in simple parametric models.

RELATED WORK

- Hausman and Newey (1995) estimate the conditional mean of gasoline demand nonparametrically
  - Their estimate is non-monotonic in price
- Blundell, Horowitz, and Parey (2012) estimate the conditional mean of gasoline demand under the Slutsky condtion.
  - Conditional mean demand may not satisfy the Slutsky condition if unobserved heterogeneity enters individual demand in a non-separable way.
  - Imposing Slutsky may lead to a misspecified model.

MORE RELATED WORK

- Hausman and Newey (2013) show that the demand function is not identified if unobserved heterogeneity is multi-dimensional.
- Hoderlein and Vanhems (2011) allow endogenous regressors in a control function approach.
- Schmalensee and Stoker (1999) estimate an Engel curve for gasoline nonparametrically but do not have price data.
- Yatchew and No (2001) estimate a partially linear model of gasoline demand.

OUTLINE

- Description of data
- Fully nonparametric estimates of demand function.
- Nonparametric estimation subject to Slutsky restriction.
- Possible endogeneity of price
- Deadweight loss of a tax
- Conclusions

DATA

- Data are from the 2001 National Household Travel Survey (NHTS).
- This is a household-level survey complemented by travel diaries and odometer readings.
- The nonparametric estimates condition on:
  - Income for the three quartiles of the income distribution.
  - Demographic and locational variables.
- The resulting sample contains 3,640 observations.

THE NONPARAMETRIC MODEL

- Notation
  - Q = Quantity demanded
  - P = Price
  - Y = Household income
  - U = Unobserved heterogeneity
- The demand function is

$$Y = g(P, Y, Q)$$

ASSUMPTIONS

- To ensure identification, we assume that
  - U is statistically independent of (P, Y).
  - $g(P, Y, U)$ is monotone increasing in U
- Given these assumptions, we assume without further loss of generality that $U \sim U[0,1]$
- Later, I consider the possibility that P is endogenous, so U is not independent of P.

NONPARAMETRIC MODEL (2)

- Under the assumptions, the α quantile of Q conditional on (P, Y, X) is

$$Q_\alpha = g(P, Y, \alpha)$$

$$\equiv G_\alpha(P, Y).$$

- Therefore, for a random variable V_α we have

$$Q = G_\alpha(P, Y) + V_\alpha; \;\; P(V_\alpha \le 0 \mid P, Y) = \alpha$$

ESTIMATION

- Estimation is based on a truncated series approximation to G_α with a B-spline basis, $\{\psi_j\}$.

- The approximation is

$$G_\alpha(P,Y) \approx \sum_{j=1}^{J_n}\sum_{k=1}^{K_n} c_{jk}\psi_j(P)\psi_k(Y)$$

- The c_{jk}'s are constants (Fourier coefficients).

- J_n and K_n are truncation points chosen by cross-validation.

ESTIMATION (2)

- The c_{jk}'s are estimated by minimizing

$$S_n(c) = \sum_{i=1}^{n} \rho \left[Q_i - \sum_{j=1}^{J_n} \sum_{k=1}^{K_n} c_{jk} \psi_j(P_i) \psi_k(Y_i) \right]$$

- ρ = check function
- $\{Q_i, P_i, Y_i : \ i = 1, \ldots, n\}$ = data

ESTIMATION UNDER SLUTSKY CONDITION

- The Slutsky condition is

$$\frac{\partial G_\alpha(P,Y)}{\partial P} + G_\alpha(P,Y)\frac{\partial G_\alpha(P,Y)}{\partial Y} \le 0$$

- Estimation consists of minimizing $S_n(c)$ subject to this constraint
- There is a continuum of constraints
- We replace the continuum with a discrete grid of values of (P,Y)

RELATION TO CONDITIONAL MEAN

- The conditional mean of demand is

$$E(Q \mid P,Y) \equiv m(P,Y) = \int g(P,Y,u)\, f_U(u)du$$

- If

$$g(P,Y,U) = m(P,Y) + U; \;\; E(U \mid P,Y) = 0,$$

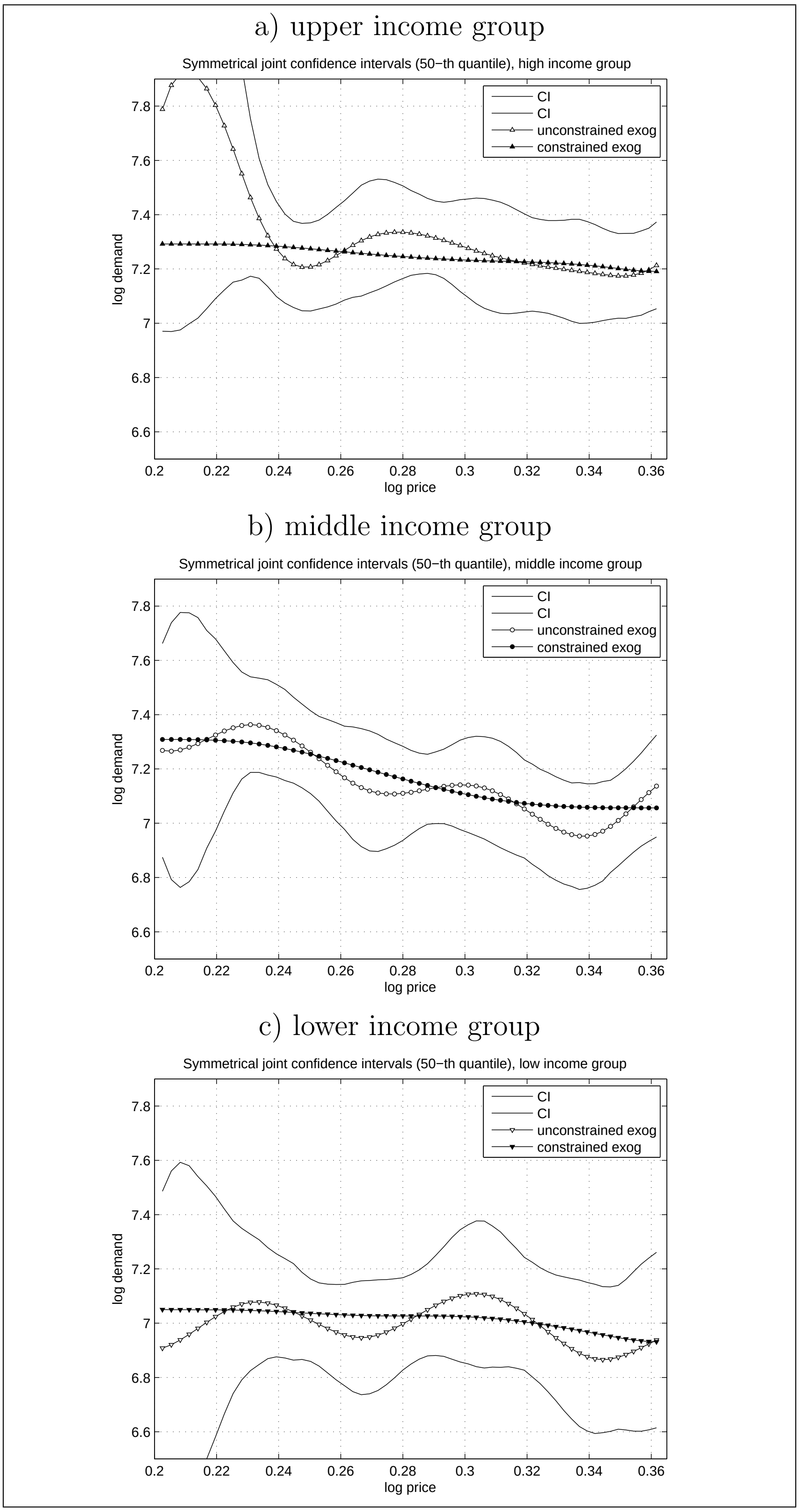
  then imposing Slutsky on $m(P,Y)$ is equivalent to imposing it on $g(P,Y,U)$ at $U = 0$.

- Otherwise, $m(P,Y)$ may not satisfy Slutsky, even if $g(P,Y,U)$ does at each U (Lewbel 2001).

MORE ON RELATION TO CONDITIONAL MEAN

- The conditional mean model imposes the Slutsky condition at only one value of U.
- The conditional quantile model imposes Slutsky at all values of U and, therefore, on all individuals.

Figure 1: Quantile regression estimates: constrained versus unconstrained estimates



Note: Figure shows unconstrained nonparametric quantile demand estimates (filled dots) and constrained nonparametric demand estimates (filled dots) at different points in the income distribution for the median ($\alpha = 0.5$), together with simultaneous confidence intervals. Income groups correspond to \$72,500, \$57,500, and \$42,500. Confidence intervals shown refer to bootstrapped symmetrical, simultaneous confidence intervals with a confidence level of 90%, based on 4,999 replications. See text for details.

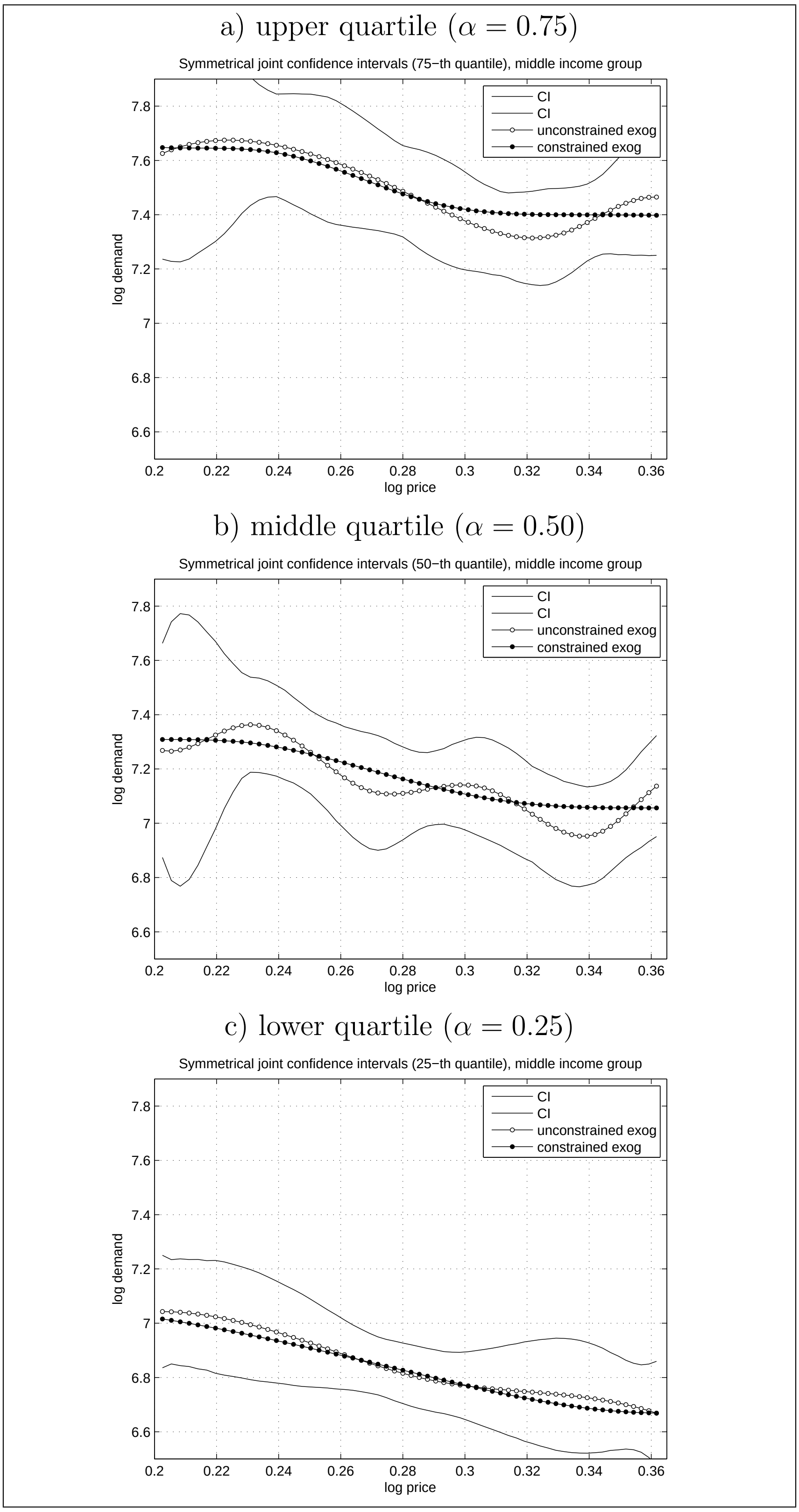
COMMENTS ON ESTIMATION RESULTS

- The nonparametric estimates are wiggly, do not satisfy the Slutsky condition, and are inconsistent with consumer theory.
- Assuming demand satisfies the Slutsky condition, wiggliness is an artifact of random sampling errors.
- The Slutsky constrained estimates are
    - Downward sloping and not wiggly.
    - Contained in 90% confidence band around unconstrained estimates

COMMENTS (2)

- The middle income group is more sensitive to price than are the outer two groups.
  - This feature of the demand function is not revealed by conventional parametric models (e.g., log-linear, log-quadratic)
- Slutsky constrained conditional mean estimates are similar to the quantile estimates.

Figure 2: Quantile regression estimates: constrained versus unconstrained estimates (middle income group)



Note: Figure shows unconstrained nonparametric quantile demand estimates (filled markers) and constrained nonparametric demand estimates (filled markers) at the quartiles for the middle income group ($57,500), together with simultaneous confidence intervals. Confidence intervals shown refer to bootstrapped symmetrical, simultaneous confidence intervals with a confidence level of 90%, based on 4,999 replications. See text for details.

COMPARISON ACROSS QUANTILES

- The constrained estimates are similar in shape and approximately parallel to one another.
- This is consistent with additive separability and homoscedasticity
  - Conditional mean estimates show shapes similar to those of the conditional quantile functions.

PRICE ENDOGENEITY

- In this model, $Q = G_\alpha(P,Y) + V_\alpha$, but $P(V_\alpha \le 0 \mid P,Y)$ is an unknown function of P.
- G_α is identified by using an instrument Z for P (distance from the Gulf of Mexico.
- The resulting model is

$$Q = G_\alpha(P,Y) + V_\alpha;\;\; P(V_\alpha \le 0 \mid Z,Y) = \alpha$$

PRICE ENDOGENEITY (2)

- Estimate G_α by solving with or without the Slutsky constraint

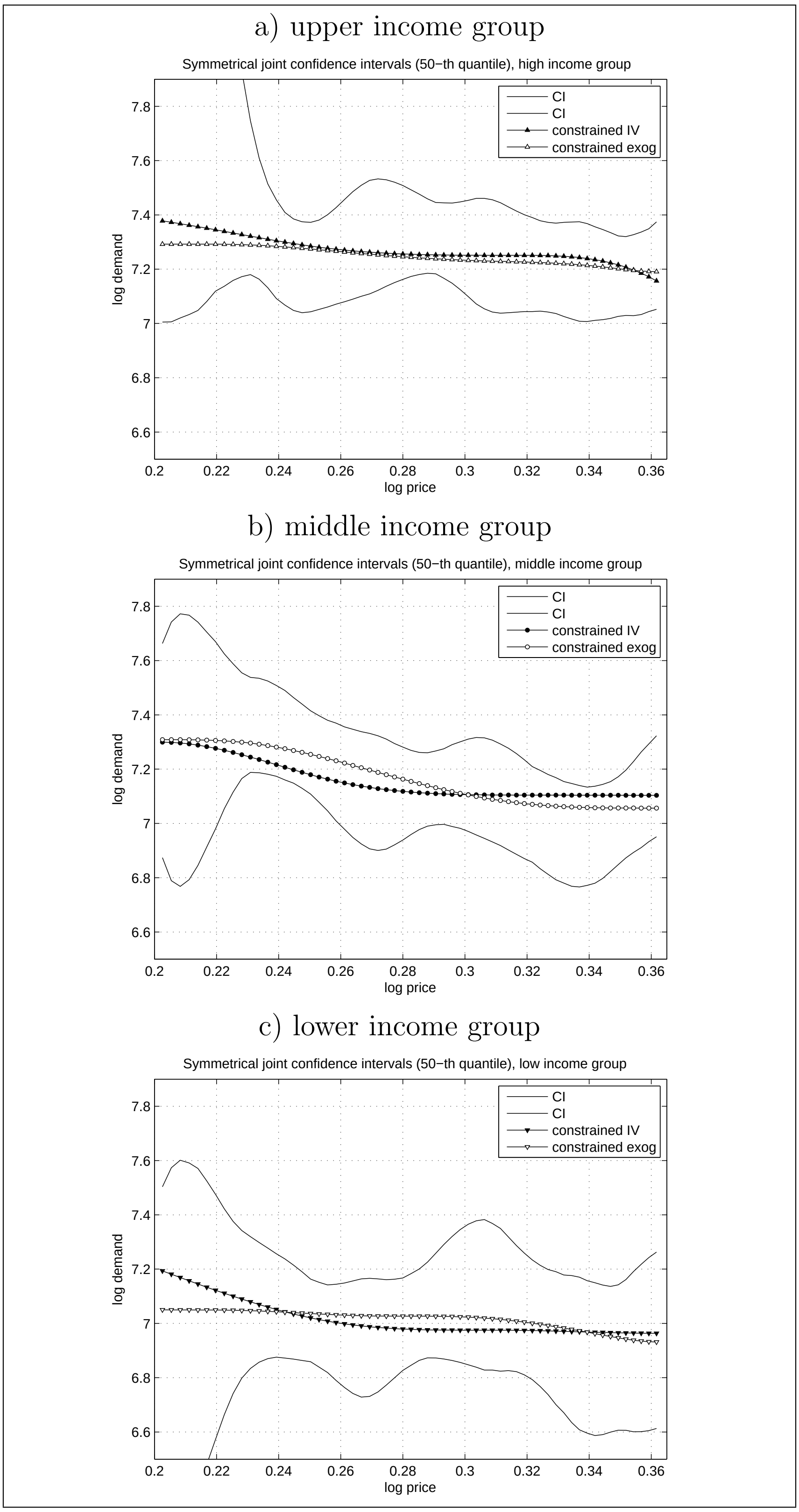
$$\underset{G_\alpha \in \mathcal{H}_n}{\text{minimize}}: \int Q_n(G_\alpha, z, y)^2 \, dzdy$$

where $\mathcal{H}_n$ is space of spline approximations and

$$Q_n(G_\alpha, z, y) =$$

$$n^{-1}\sum_{i=1}^{n}\{I[Q_i - G_\alpha(P_i, Y_i) \leq 0] - \alpha\} I(Z_i \leq z; Y_i \leq y)$$

Figure 6: Quantile regression estimates under the shape restriction: IV estimates versus estimates assuming exogeneity



Note: Figure shows constrained nonparametric IV quantile demand estimates (filled markers) and constrained quantile demand estimates under exogeneity (open markers) at different points in the income distribution for the median ($\alpha = 0.5$), together with simultaneous confidence intervals. Income groups correspond to \$72,500, \$57,500, and \$42,500. Confidence intervals shown correspond to the unconstrained quantile estimates under exogeneity as in Figure 1. See text for details.

DEADWEIGHT LOSS

- Estimate deadweight loss of a tax by integrating demand function to obtain expenditure function.
- Assumed tax changes price from 5th to 95th percentile of price in sample.
- Some estimates of deadweight losses using unconstrained demand function are negative.
    - This is unsurprising given non-monotonicity of unconstrained estimated demand function.
    - Constrained estimates have correct signs and show that middle income group has the largest loss.

CONCLUSIONS

- Nonparametric estimates of demand functions eliminate risk of specification error but can be poorly behaved due to random sampling errors.
- Constraining nonparametric estimates to satisfy the Slutsky condition overcomes this problem without need for arbitrary parametric or semiparametric restrictions.
- In a non-separable model of gasoline demand
    - Fully nonparametric estimates are non-monotonic
    - Constrained estimates are monotonic and reveal features not easily found with parametric models.